Mapping the Gene Space with **Gene Signal Pattern Analysis**

Aarthi Venkat[1], Martina Damo[2], Nikhil S. Joshi[2], Smita Krishnaswamy[1]

[1] Computational Biology and Bioinformatics Program, Yale University

[2] Department of Immunobiology, Yale School of Medicine

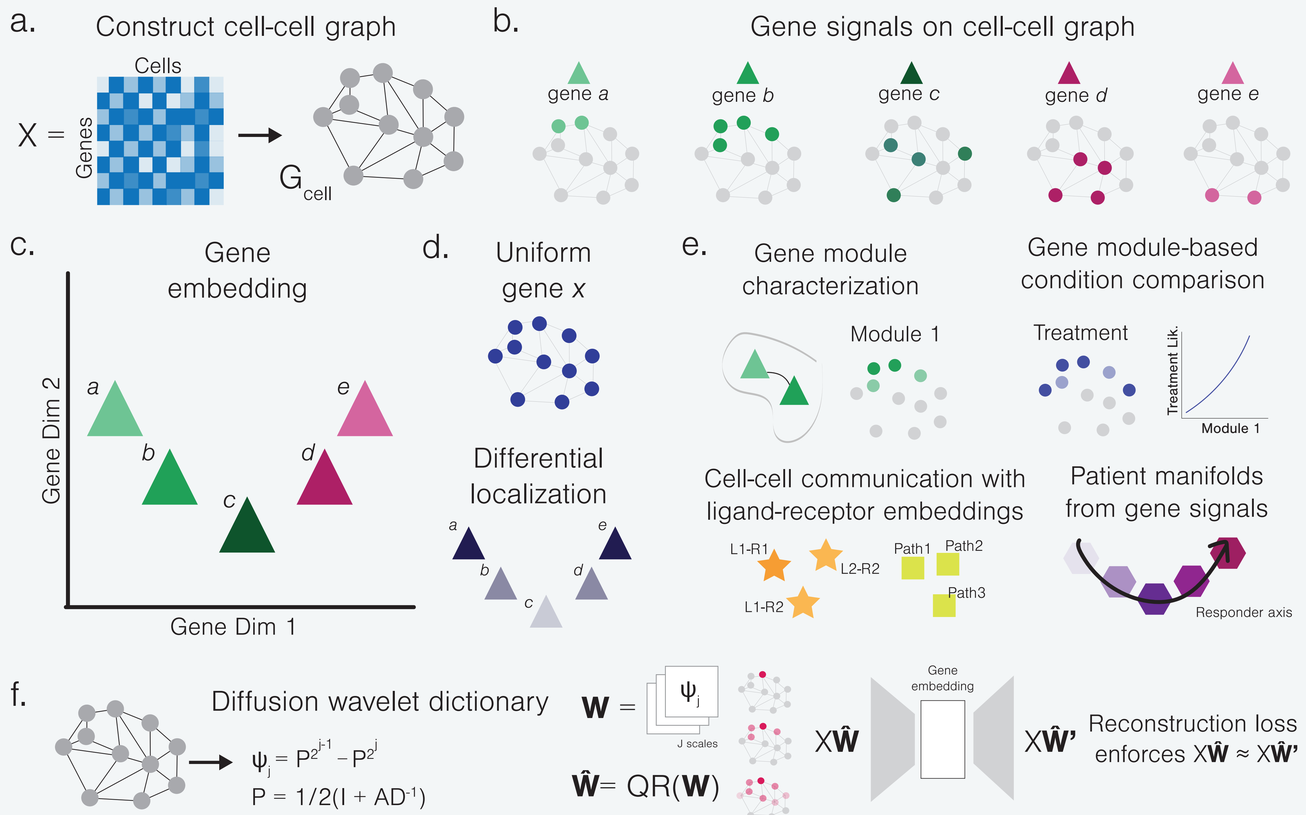
Abstract

In high-dimensional single-cell analysis, several computational methods have been developed to map the cellular state space, but little has been done to map the gene space. A mapping that preserves gene-gene relationships within the dataset is particularly useful for characterizing cellular heterogeneity within cell types, where boundaries between cell subpopulations are often unclear or even arbitrary.

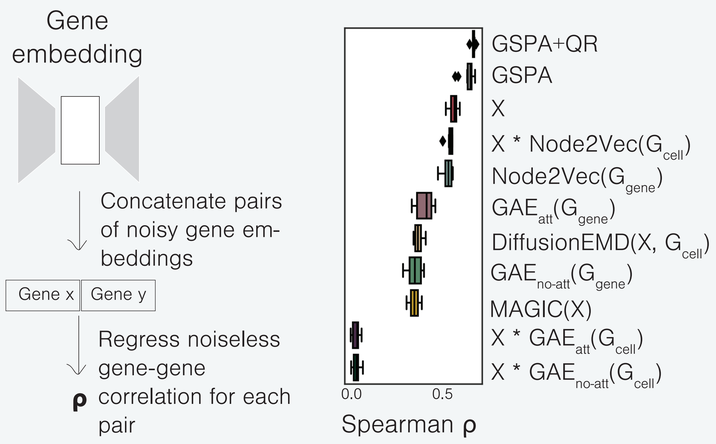
Here, we present gene signal pattern analysis, a new paradigm for analyzing single cells. We build a cell-cell graph and design a dictionary of diffusion wavelets, capturing a multiscale view of the cell space. We then transform genes by the dictionary and learn a reduced gene representation. Given the gap in prior research for this problem, we design nine alternative strategies and three benchmarks for evaluating preservation of gene-gene relationships, all of which are outperformed by diffusion wavelet-transformed signals. We also define, calculate, and evaluate localization, a key property of a gene signal on the cellular graph.

We demonstrate the utility of gene signal pattern analysis (GSPA) on T cells from a mouse model of peripheral tolerance in skin[1]. GSPA reveals a continuum of gene signals characterized by T cell subtypes and transcriptional programs related to effector function and proliferation. In the same model, GSPA captures the key groups of ligand-receptor pairs with shared patterns, including PD-L1/PD-1 communication between a subset of myeloid and T cells. Finally, we built a multiscale manifold of 48 melanoma patient samples, demonstrating the ability of our method to characterize differences between responders and non-responders to checkpoint immunotherapy[2]. Together, we show gene signal pattern analysis, through methodology from graph signal processing, spectral graph theory, and machine learning, represents an avenue for future research in scRNA-seq analysis.

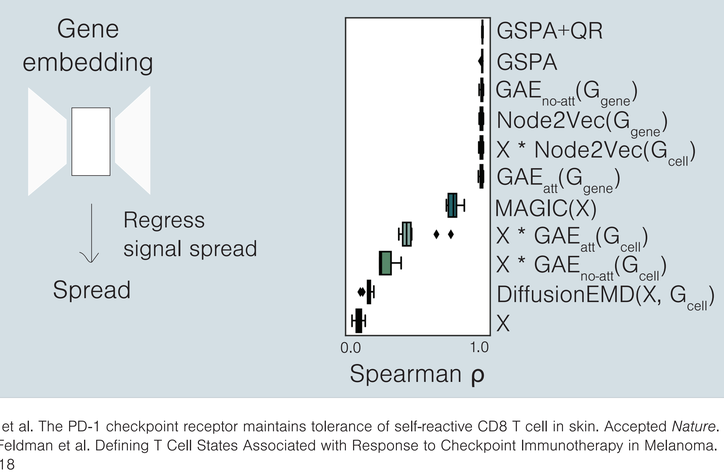
Overview of Gene Signal Pattern Analysis (GSPA)



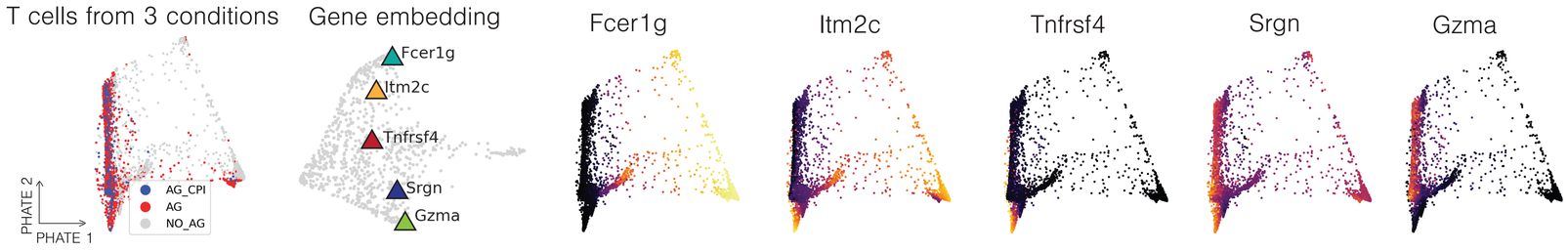
Predict gene-gene correlation from Splatter-simulated data



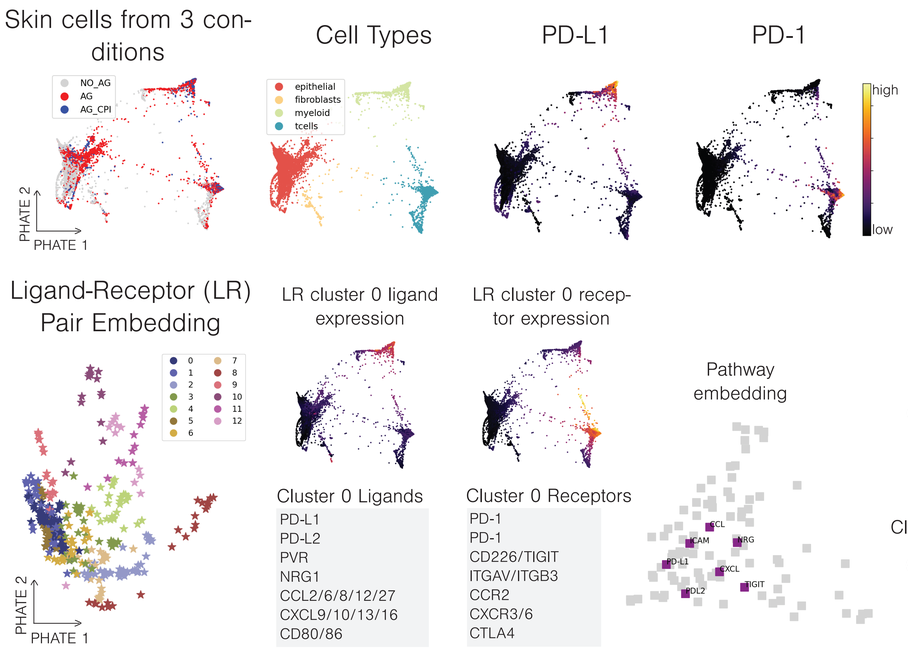
Predict gene localization



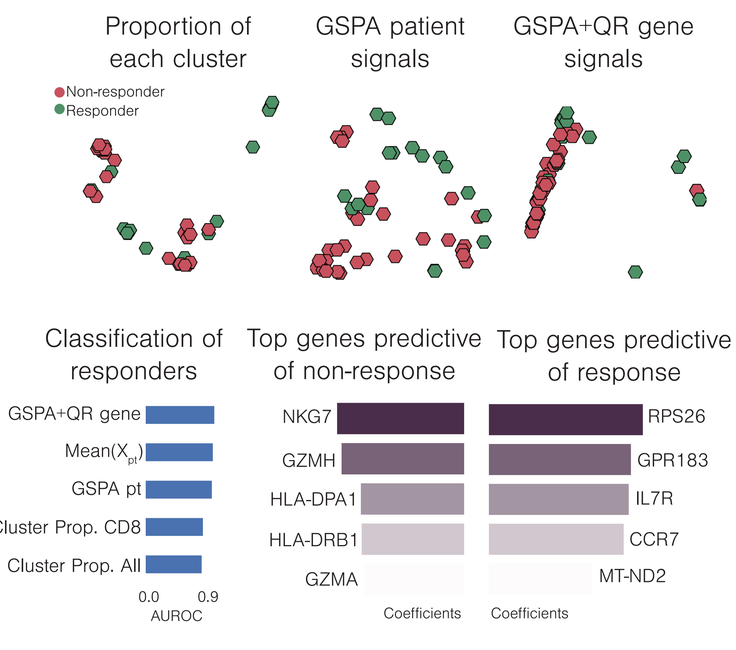
Gene embeddings reveal complex gene patterns



Cluster-independent cell-cell communication



Patient manifolds



[1] Damo et al. The PD-1 checkpoint receptor maintains tolerance of self-reactive CD8 T cell in skin. Accepted *Nature*.
[2] Sade-Feldman et al. Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell* 2018